

Web search: public searching of the Web (A. Spink; B. J. Jansen)

Por Aurora González Teruel

EL IMPACTO Y CRECIMIENTO DE LA WEB en diferentes contextos ha motivado su estudio, fundamentalmente, desde un punto de vista tecnológico. No obstante, no podemos olvidar que al otro lado de esa plataforma existen individuos con unas necesidades informativas, unas motivaciones y unos condicionantes particulares. Precisamente, el libro *Web search: public searching of the Web* asume esa visión; concretamente se dirige a establecer un marco de referencia para el estudio de los motores de búsqueda en internet, pero no sólo de sus capacidades, avances o eficacia de recuperación, sino también de los hábitos de búsqueda de los usuarios de esos servicios.

De esta manera Spink y Jansen ofrecen una perspectiva amplia de la investigación existente sobre la búsqueda de información en la

web con especial atención a la serie de estudios realizados por los propios autores durante los años 1997 a 2004 a partir del análisis de los logs de una serie de buscadores de internet y que previamente fueron publicados en forma de artículo de revista.

Amanda Spink es actualmente *associate professor* en la *School of*

Spink, Amanda; Jansen, Bernard J.

Web search: public searching of the Web. Boston: Kluwer Academic Publishers, 2004. ISBN: 1-4020-2268-9.

Information Sciences de la *University of Pittsburgh*. Sus líneas de investigación incluyen el estudio de la conducta informativa y la recuperación de información desde un punto de vista cognitivo. Por otra parte, Bernard J. Jansen es *assistant professor* en la *School of Information Sciences and Technology* de

la *Pennsylvania State University* y sus investigaciones se han centrado, entre otros aspectos, en la recuperación de información en la web y en el diseño de sistemas de información. Por lo tanto, el bagaje de ambos investigadores ofrece una visión complementaria: el punto de vista del usuario y el del sistema utilizado por éste para buscar información en la web.

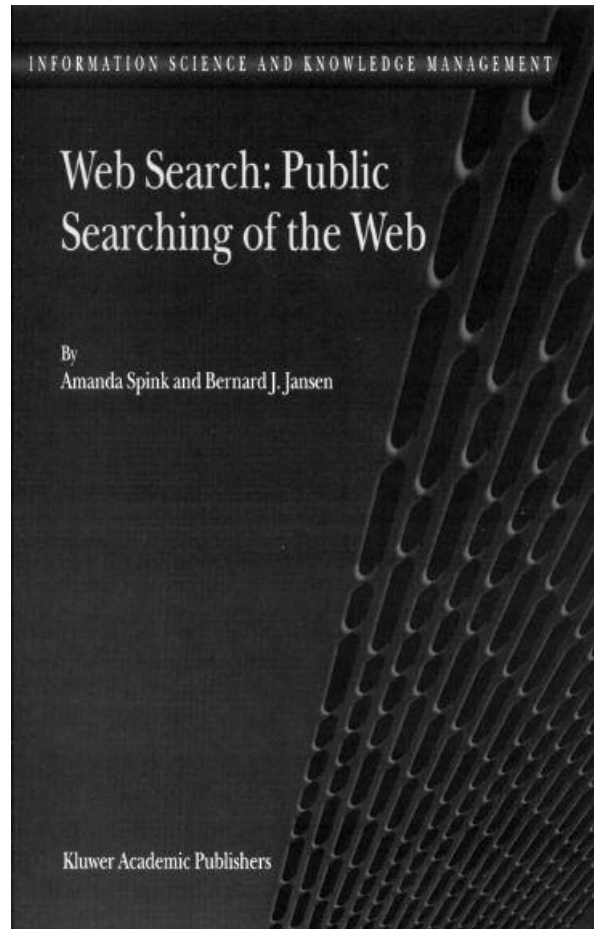
El libro, según afirman los propios autores, se dirige tanto a investigadores como a estudiantes de documentación, de informática y de diversas ciencias cognitivas. Pero igualmente está pensado para consultores web, webmasters y proveedores de contenidos. En este último caso, la valoración que presenta respecto a los hábitos de búsqueda de información de distinta naturaleza supone una información de gran valor desde el punto de vista de la planificación de estos servicios.

El libro está estructurado en cuatro secciones. En la primera, *The context of web search* (capítulos 1 al 3), se establece el marco teórico para estructurar la investigación existente sobre búsqueda en web planteando la existencia de los siguientes niveles de análisis: el contexto tecnológico, el social, el organizacional, el de la conducta informativa humana y el de la interacción hombre-ordenador en la búsqueda en web. En el capítulo uno se tratan los tres primeros casos. En primer lugar, en el contexto tecnológico se establece una panorámica de los motores de búsqueda existentes en internet, los fundamentos de su funcionamiento y una breve aproximación cuantitativa al tamaño de la web y a la actividad de búsqueda de información que soporta. Posteriormente se aborda el estudio de la búsqueda en la web desde el punto de vista del contexto social, especialmente con relación al uso doméstico que se hace de este recurso. Para ello, se realiza una revisión bibliográfica de una serie de trabajos que investigan la web desde esta perspectiva. Por último, se analiza el contexto organizacional en el estudio de la búsqueda en web, mencionando únicamente dos monografías que asumen esta perspectiva. En general, aunque el planteamiento de este marco teórico tiene valor por sí mismo, el desarrollo desigual de los distintos niveles que conciben está ciertamente descompensado.

El capítulo dos está dedicado a describir brevemente los otros dos niveles contemplados en el marco teórico del estudio de la búsqueda en la web planteados en el capítulo

anterior: el de la conducta informativa humana y el de la interacción persona-ordenador. Además, mencionan algunos métodos utilizados en el uso de la web y revisan varios estudios realizados desde 1995, clasificados como estudios sobre:

—comportamiento en la búsqueda en la web; es decir, por qué y cómo las personas buscan en este entorno,



—búsqueda en un único sitio,

—búsqueda en la web desde el punto de vista de la *Optimal Foraging Theory (OFT)*,

—comportamiento en la búsqueda en la web por parte de los niños,

—formación para la búsqueda en la web y

—evaluación de la interacción con buscadores tanto desde el punto de vista de la usabilidad como de

la eficacia de estas herramientas de consulta.

En el capítulo tres se exponen diversos aspectos del diseño de la investigación realizada por los autores. Concretamente, se define y discute la técnica utilizada para la recogida de datos, el análisis del registro de transacciones (*transaction log analysis*). Además, se expone la estructura y características de los logs de cada uno de los buscadores analizados (*AltaVista, Excite, Allthe-web*). En tercer lugar, se detallan los aspectos analizados tanto desde el punto de vista cuantitativo como cualitativo. Por último, se describen las limitaciones de un estudio de esta naturaleza.

En la sección II, *How people search the web*, se exponen detalladamente los resultados de los investigaciones antes mencionadas. En los tres capítulos que la componen se muestran los datos a nivel de términos de búsqueda, de estrategia de consulta y de sesiones de búsqueda. El análisis desde el punto de vista de los términos utilizados para la búsqueda en web (capítulo 4) contiene los más frecuentemente utilizados, la co-ocurrencia de términos y las materias buscadas más habitualmente. En segundo lugar, los resultados desde el punto de vista de las estrategias de búsqueda (capítulo 5) presentan los datos sobre el tamaño de estas estrategias, el uso de las utilidades avanzadas, las estrategias repetidas, la preferencia de idioma y los documentos web vistos por estrategia de búsqueda aplicada. Por último, en el análisis de las sesiones de búsqueda (capítulo 6) se muestran los resultados sobre la longitud de la sesión de

búsqueda, la duración, las páginas de resultados vistas, documentos por sesión, relevancia de los documentos vistos.

Los resultados expuestos en el punto anterior se detallan en la sección III (*Subjects of web search*) con un análisis de las pautas de búsqueda en la web en cuatro áreas: comercio electrónico (capítulo 7), medicina y salud (capítulo 8), sexo (capítulo 9) y multimedia (capítulo 10). Estos resultados proporcionan información de gran valor para la adecuación, mejora o el diseño de nuevas herramientas específicas de búsqueda en la web. En este sentido es interesante conocer las dificultades con las que se encuentran los usuarios cuando buscan información sobre comercio electrónico, dado el volumen de la información disponible. O, por ejemplo, las limitaciones de los buscadores gene-

ralistas para la búsqueda de información en salud cuando se utiliza un vocabulario tan especializado como el médico debido, en gran medida, al hecho de que estas herramientas fueron concebidas inicialmente para facilitar el uso del lenguaje natural.

Por último, en la sección IV (*Key findings, trends, further research and conclusions*) se exponen las conclusiones de los distintos estudios que los autores han realizado así como el camino que deben emprender futuras investigaciones.

En un contexto tan cambiante como el que se describe, esta obra supone un valioso esfuerzo de síntesis y revisión que, planteando un marco teórico y metodológico, abre las puertas a futuras investigaciones. No obstante, precisamente por ese contexto cambiante y no consolidado en el que se centra, no puede

considerarse un texto definitivo sino una referencia importante para investigadores y estudiosos de los hábitos de búsqueda de información en la web. En este sentido, la exposición de la metodología de los estudios llevados a cabo por ambos autores ofrece una importante información para el diseño investigaciones similares. Por otra parte, la presentación de sus resultados, independientemente de la omisión de buscadores tan utilizados como *Google*, supone una fuente de datos de gran valor de cara a la planificación de herramientas de búsqueda en internet o al posicionamiento en este medio de sitios web de diversa naturaleza.

Reseña enviada a Epi por:
Aurora González Teruel, Instituto de Historia de la Ciencia y Documentación López Piñero (CSIC-UVEG).
aurora.gonzalez@uv.es